行政院國家科學委員會專題研究計劃成果報告
單階段和雙階段與平均之多重比較程序的模擬比較

A Simulation Comparison of Single-Stage and Two-Stage Multiple Comparison Procedures With the Average

計劃編號: NSC90-2118-M-032-011

執行期限: 90 年 8 月 1 日至 91 年 7 月 31 日

主持人:吳淑妃淡江大學統計系

October 1, 2002

中文摘要

在許多實驗中我們常常將所有實驗對象的平均值當作一個評量的標準，希望能夠從中找出比整體平均好，比整體平均差，或是與整體平均無明顯差異的子集合。例如在農業上可用來篩選出比平均優良的小麥品種，使種植時的效益更高；或是在醫學臨床上，可針對具有相同療效的藥品中，找出比平均效果更好的。

本文中，當變異數未知且可能不相等時，我們採用了三種不同的方法來求取常態分配平均數與平均之聯立信賴區間，分別是傳統法、單階段和雙階段與平均之多重比較程序。並使用蒙地卡羅模擬法來模擬出傳統法，單階段和雙階段與平均之多重比較程序的信賴區間長度及信心水準。由模擬結果發現發現單階段和雙階段與平均之多重比較程序之模擬信心水準皆可達到名目信心水準。並舉一個生統的例子以示範單階段，雙階段與平均之多重比較程序。

關鍵辭:

單階段與平均之多重比較程序, 雙階段與平均之多重比較程序.

ABSTRACT

In many experimental situations, the average treatment performance within its own group is used as a benchmark to be compared with each individual treatment. Our study propose is to identify better than the average, worse than the average and not much difference from the average subsets based on the simultaneous two-sided confidence interval of each normal mean from its average. In this article, a simulation study of traditional, single-stage and two-stage multiple comparison procedures with the average for normal distribution under heteroscedasticity is investigated by the Monte-Carlo techniques. The length of simultaneous confidence interval and confidence coefficient for three procedures are simulated and it's found that the simulated confidence coefficients can reach its nominal confidence coefficients for single-stage and two-stage multiple comparison procedures with the average while the traditional procedure fails to under heteroscedasticity. A biometrical example is given to illustrate the single-stage and two-stage procedures.

1. 計劃緣由與目的

令 $\pi_1, ..., \pi_k$ 為 k 個互相獨立之常態分配母體，表為 $N(\theta_i, \sigma_i^2)(i = 1, ..., k)$，其中 $\theta_1, ..., \theta_k$ 為未知的平均數，$\sigma_1^2, ..., \sigma_k^2$ 為 k 個變異數，令 k 個平均數之平均為 $\bar{\theta} = \frac{\sum_{i=1}^k \theta_i}{k}$。對於 $\theta_i - \bar{\theta}(i = 1, ..., k)$ 的 $(1-\alpha)100\%$ 雙邊聯立信賴區間，當變異數相等時，Wu and Chen(1998) 提出了常態分配平均數與平均的多重比較程序，給定為 $A_i = (\bar{X}_i - \bar{\bar{X}} - c\sqrt{\sigma_{ii}}, \bar{X}_i - \bar{\bar{X}} + c\sqrt{\sigma_{ii}}), i = 1, ..., k$，其中 $\bar{X}_i$ 為母體 π_i 的 n_i 個獨立觀測值的樣本平均數 $(i = 1, ..., k)$，$\bar{\bar{X}} = \sum_{i=1}^k \bar{X}_i/k$，而 $\sigma_{ij} = \sigma^2 d_{ij}$，其中

$$d_{ij} = \begin{cases} \frac{(k-1)^2}{k^2}\frac{1}{n_i} + \frac{1}{k^2}\sum_{l\neq i}\frac{1}{n_l}, & i = j \\ -\frac{(k-1)}{k^2}(\frac{1}{n_i} + \frac{1}{n_j}) + \frac{1}{k^2}\sum_{l\neq i,j}\frac{1}{n_l}, & i \neq j \end{cases}$$

當樣本數相等時，利用 Nelson(1993) 的程式得到聯立信賴區間的精確臨界值 c；當樣本數不等且 $k = 3$ 時，他們利用 "structure l" 可得到精確臨界值，當 $k > 3$ 時，則只可得到近似臨界值。當變異數未知且可能不等時，他們則提出了單階段與平均比較的多重比較程序。單階段抽樣程序是令 $X_{ij}(j = 1, ..., n_i)$ 為從常態母體 $\pi_i(i = 1, ..., k)$ 中取出的 n_i 個獨立隨機樣本，將 n_i 個樣本分成兩部分，第一部分為 n_0 個樣本，第二部分則為 $n_i - n_0$ 個樣本，利用第一部分的樣本，其樣本平均和樣本變異數可得加權統計量

$$\tilde{X}_i = \sum_{j=1}^{n_i} W_{ij}X_{ij},$$

其中權數 W_{ij} 定義為

$$W_{ij} = \begin{cases} U_i, 1 \leq j \leq n_0 \\ V_i, n_0 + 1 \leq j \leq n_i \end{cases},$$

$$U_i = \frac{1}{n_i} + \frac{1}{n_i}\sqrt{\frac{n_i - n_0}{n_0}(\frac{n_i c^{*^2}}{S_i^2} - 1)},$$

$$V_i = \frac{1}{n_i} - \frac{1}{n_i}\sqrt{\frac{n_0}{n_i - n_0}(\frac{n_i c^{*^2}}{S_i^2} - 1)},$$

其中 $c^{*^2} = \max_{1\leq j\leq k}\{\frac{S_j^2}{n_j}\}$，則雙邊聯立信賴區間給定為 $(\tilde{X}_i - \bar{\tilde{X}} \pm c^* h_t^*), i = 1, ..., k$，其中 $\bar{\tilde{X}} = \sum_{i=1}^k \tilde{X}_i/k$。Wen and Chen(1994) 證明下列轉換

$$Y_i = \frac{\tilde{X}_i - \theta_i}{\sqrt{S_i^2 \sum_{j=1}^{n_i} W_{ij}^2}} = \frac{\tilde{X}_i - \theta_i}{c^*}, i = 1, ..., k,$$

為自由度 $n_0 - 1$ 的獨立 t 分配。為了得到更精確的臨界值，則 h_t^* 修改後的模擬法如下

$$\begin{aligned} & P \quad (-h_t^* \leq V_{[1]} \leq V_{[k]} \leq h_t^*) \\ & = \quad P(|V|_{[k]} \leq h_t^*), \end{aligned}$$

其中 $V_i = Y_i - \bar{Y}, \bar{Y} = \frac{\sum_{i=1}^k Y_i}{k}$，$|V|_{[1]} \leq |V|_{[2]} \leq \cdots \leq |V|_{[k]}$ 為 V_i 取絕對值後再排序之統計量。上面等式之所以成立是因為 V_i 的分配對稱於原點。令上式等於 P^*，則可求出 h_t^* 之精確值，h_t^* 的 Monte-Carlo 模擬結果已列表。Wu and Chen(2000) 則使用雙階段抽樣方法來解決變異數不相等時，常態分配平均數與平均的多重比較問題。而雙階段抽樣程序是先從母體 π_i 中取出 $X_{i1}, X_{i2}, ..., X_{in_0}$ 共 $n_0(\geq 2)$ 個起始樣本，然後根據此 n_0 個樣本，計算樣本變異數 S_i^2，其為 σ_i^2 的不偏估計量。則第 i 個母體之最後樣本總數 n_i 定義如下

$$n_i = max\{n_0 + 1, [\frac{S_i^2}{c^2}] + 1\}, \qquad (1)$$

$[x]$ 為小於 x 的最大整數。即對第 i 個母體，尚須多取 $n_i - n_0$ 個額外的隨機

觀察值，以使總觀測值有n_i個，表為$X_{i1},...,X_{in_0},...,X_{in_i}$。則雙邊聯立信賴區間給定為$(\tilde{X}_i-\bar{\tilde{X}}\pm ch_t)$，其中

$$\tilde{X}_i=\sum_{j=1}^{n_i}a_{ij}X_{ij}, \tag{2}$$

其中係數$a_{i1},...,a_{in_0},...,a_{in_i}$決定如下：

$$\begin{aligned} a_{i1} &= \ldots = a_{in_0} = \frac{1-(n_i-n_0)b_i}{n_0} = a_i, \\ a_{i,n_0+1} &= \ldots = a_{in_i} = \frac{1}{n_i}[1+\sqrt{\frac{n_0(n_ic^2-S_i^2)}{(n_i-n_0)S_i^2}}] = b_i. \end{aligned}$$

。因為雙階段的T_i和單階段的Y_i有相同的分配，所以臨界值h_t和雙階段的臨界值h_t^*相同，其模擬的精確值皆和單階段的相同。

我們的模擬研究利用陳順益(1998)在其表2中的結構以及Monte-Carlo模擬法來比較傳統法，單階段與雙階段與平均比較之聯立信賴區間之信心水準和信賴區間長度上的表現差異。我們考慮在四個母體的情況下，四種不同的樣本數組合 ：(6,6,6,6),(6,6,8,10),(12,12,12,12) 及(6,10,16,20)，並分別考慮變異數相等$(\sigma_1^2=1,\sigma_2^2=1,\sigma_3^2=1,\sigma_4^2=1)$和變異數不等$(\sigma_1^2=1,\sigma_2^2=4,\sigma_3^2=4,\sigma_4^2=9)$的情形。首先產生服從常態分配的亂數，並使其具有我們所控制的變異數結構，然後利用這些模擬產生的樣本來做模擬比較，在傳統法中，根據第三章第一節中變異數相等但未知時，計算出其聯立信賴區間$(\bar{X}_i-\bar{\bar{X}}-cS_p\sqrt{d_{ii}},\bar{X}_i-\bar{\bar{X}}+cS_p\sqrt{d_{ii}}),i=1,2,3,4$，並檢驗$\theta_i-\bar{\theta}$是否有落在所屬的信賴區間之中，$\theta_i$可為任意值，如此重複模擬30000次，則模擬的信心水準為$\theta_i-\bar{\theta}$落在所屬信賴區間中的百分比。在雙階段抽樣程序中，設第一階段所用的樣本數為$n_0=10$，利用第三章第二節的方法，控制信賴區間的長度與傳統法相同，則可決定總樣本數n_i，產生n_i個常態分配的隨機變數，具有我們所控制的變異數結構，並求出聯立信賴區間$(\tilde{X}_i-\bar{\tilde{X}}\pm ch_t),i=1,2,3,4$，如此模擬30000次，則模擬的信心水準為$\theta_i-\bar{\theta}$落在所屬信賴區間中的百分比，並可求出雙階段抽樣程序和傳統法總樣本數的比值。而在單階段的抽樣程序中，根據Stein(1945)提出的方法採用雙階段抽樣方法期望樣本數的上界值m_i，

$$\begin{aligned} m_i &= (n_0+1)P(\chi_{n_0-1}^2<\frac{(n_0^2-1)c^2}{\sigma_i^2})+\frac{\sigma_i^2}{c^2}P(\chi_{n_0-1}^2>\frac{(n_0^2-1)c^2}{\sigma_i^2}) \\ &+ P(\chi_{n_0-1}^2>\frac{(n_0^2-1)c^2}{\sigma_i^2}), \end{aligned}$$

則總樣本數決定如下：

$$n_i=[m_i],$$

此處$[x]$是不大於x的最大整數。在單階段抽樣程序中，其初始樣本設為$n_0=10$，以傳統法和雙階段相同的方法模擬出實際的信心水準、聯立信賴區間長度以及樣本比值做比較。

2.結果與討論

其模擬的結果顯示當變異數相等時，三種方法其模擬信心水準皆能夠達到名目信心水準，可是當變異數不等時，使用傳統法會發生模擬信心水準無法達到名目信心水準的情形，若就單階段和雙階段作比較，會發現雙階段法之聯立信賴區間的長度大部分皆比單階段窄，而其所需的總樣本數皆大致相同。另外我們發現當變異數不相等時，傳統法的聯立信賴區間明顯變寬，若是使用雙階段抽樣程序則可調整總樣本數來控制信賴區間長度，若要和傳統法有相同的信賴區間長度，則所需抽樣的樣本數由1.44倍至2.04倍。此外我們還對大樣本的情形做模擬，此時的變異數皆令為1且只考慮等樣本大小n的情形，雙階段抽樣程序的第一階段樣本和單階段抽樣程序的初始樣本皆定為20，對母體數目$k=6,8,10(10)50,70,100$，$n=10,20,30,60,100$進行模擬的結果可看出不論母體數是多少，當總樣本數增加時，均能使聯立信賴區間的長度減少。而雙階段和單階段的樣本總數大致上相等，我們可由表中與傳統法相比的樣本數比值看出，隨著n的增加，雙階段和單階段的樣本數

比例也逐漸降低。在表中我們把雙階段的信賴區間長度控制為與傳統法相同，而雙階段的聯立信賴區間大多數是比單階段短，除了 $(k,n) = (6,10),(8,10),(10,10),(20,10)$ 當 $P^* = .90$, $(k,n) = (6,10),(8,10),(10,10)$ 當 $P^* = .95, .99$ 時。

3. 計劃成果自評

單階段和雙階段與平均之多重比較程序模擬的精確臨界值皆已完成C++程式. 當 P^* 分別為 .90, .95, .99 時, 皆已列表完成. 由模擬比較結果，我們可知當各個母體的變異數不相等時，若仍然使用傳統的方法來求聯立信賴區間，則其無法達到名目信心水準。而雙階段抽樣方法雖然可以達到名目的信心水準而且可控制信賴區間的長度，但是所需的總樣本數並無法在事前決定，故實驗的花費是未知的，在經費有限的情況下，可能會發生超出預算的情形。而單階段抽樣方法則沒有上述雙階段抽樣方法的缺點，而且模擬的信心水準也能達到名目的信心水準。故在使用這兩種方法時，可根據實際需要情形來選擇使用哪種方法較合適。如果實驗的經費足夠，且額外樣本的取得容易，則使用雙階段抽樣方法較合適。若只有有限的經費或是只有一定數目的樣本，無法再取得額外的樣本，則可使用單階段抽樣方法。關於Bishop and Dudewicz(1978)之生統例子, 因其變異數被檢定為有顯著不等, 單階段與雙階段與平均比較之多重比較程序的才可被應用。

表 1：四種殺菌劑90%雙階段聯立信賴區間

	參數	$(\tilde{X}_i - \tilde{X} - ch_t, \tilde{X}_i - \tilde{X} + ch_t)$
1	$\theta_1 - \theta$	(0.712,1.882)
2	$\theta_2 - \bar{\theta}$	(-1.614,-0.444)
3	$\theta_3 - \bar{\theta}$	(-2.077,-0.2695)
4	$\theta_4 - \bar{\theta}$	(0.638,1.808)

表 2：四種殺菌劑95%,99%雙階段聯立信賴區間

	參數	$(\tilde{X}_i - \tilde{X} - ch_t, \tilde{X}_i - \tilde{X} + ch_t)$	
		95%	99%
1	$\theta_1 - \bar{\theta}$	(0.623,1.971)	(0.432,2.162)
2	$\theta_2 - \bar{\theta}$	(-1.703,-0.355)	(-1.894,-0.164)
3	$\theta_3 - \bar{\theta}$	(-2.166,-0.818)	(-2.357,-0.627)
4	$\theta_4 - \bar{\theta}$	(0.549,1.897)	(0.358,2.088)

表 3：四種殺菌劑90%,95%,99%單階段聯立信賴區間

	參數	$(\tilde{X}_i - \tilde{X} - c^* h_t, \tilde{X}_i - \tilde{X} + c^* h_t)$		
		90%	95%	99%
1	$\theta_1 - \bar{\theta}$	(0.5184,1.6844)	(0.4304,1.7724)	(0.2394,1.
2	$\theta_2 - \bar{\theta}$	(-1.6603,-0.4063)	(-1.7483,-0.4063)	(-1.9393,-0
3	$\theta_3 - \bar{\theta}$	(-2.1085,-0.9425)	(-2.1965,-0.8545)	(-2.3875,-0
4	$\theta_4 - \bar{\theta}$	(0.9184,2.0844)	(0.8304,2.1724)	(0.6394,2.

由表1-3的結果可看出第一種和第四種殺菌劑的效果比平均還要好, 而第二和第三種比平均值差。且不論 P^* 為何值，單階段聯立信賴區間皆比雙階段來的略窄。所以此研究之程式和模擬部份也已完成. 成果和我們預期的也很一致.

4. 參考文獻

中文部份

陳順益(1998).「變異數不相等時的單階段變異數分析法」，中國統計學報, 36(4), 321-338.

英文部份

Bishop, T. A. and Dudewicz, E. J. (1978). Exact analysis of variance with unequal variances: test procedures and tables. *Technometrics*, 20, 419-430.

Chen, H. J. and Dudewicz, E. L.(1976). Procedures for fixed-width interval estimation of the largest normal mean. *Journal of American Statistical Association*, 71, 752-756.

Chen, H. J. and Lam, K.(1989). Single-stage interval estimation of the largest normal mean under heteroscedasticity. *Communications in Statistics - Theory and Methods.*, 18(10), 3703-3718.

Dudewicz,E. J. and Dalal, S. R.(1975). Allocation of observations in ranking and selection with unequal variances. *Sankhya, Ser. B.*, 37, 28-78.

Halperin,M.,Greenhouse,S.W.,Cornfield,J.and Zalokar,J.(1955). Tables of percentage points for the studentized maximum absolute deviate in normal samples. *Journal of American Statistical Association*, 50, 185-195.

Nelson, L. S.(1983). Exact critical values for use with the analysis of means. *Journal of Quality Technology*, 15, 40-44.

Nelson, P. R.(1993). Additional uses for the analysis of means and extended tables of critical values. *Technometrics*, 35, 61-71.

Nelson, P. R.(1982). Exact critical points for the analysis of means. *Communications in Statistics-Theory and Methods*, 11, 699-709.

Stein. C. M. (1945). A two-sample test for a linear hypothesis whose power is independent of variance. *Annals of Mathematical Statistics.*, 16, 243-258.

Tong. Y. L. (1980). Probability inequalities in multivariate distribution. *Academic Press*, New York.

Wen, M. J. and Chen, H. J.(1994). Single-stage multiple comparison procedure under heteroscedasticity. *American Journal of Mathematical and Management Sciences*, 14, Nos 1 and 2, 1-48.

Wu, S. F. and Chen, H. J.(2000). Two-stage multiple comparisons with the average for normal distributions under heteroscedasticity. *Computational statistics & data analysis*, 33, 201-213.

Wu, S. F. and Chen, H. J.(1998). Multiple comparisons with the average for normal distributions. *American Journal of Mathematical and Management Sciences*, 18, 193-218.